EXAMPLAR-BASED INPAINTING BASED ON LOCAL GEOMETRY

Olivier Le Meur, Josselin Gautier, Christine Guillemot

EXAMPLAR-BASED INPAINTING BASED ON LOCAL GEOMETRY

Olivier Le Meur, Josselin Gautier

University of Rennes 1
Campus Universitaire de Beaulieu
35042 Rennes Cedex France

Christine Guillemot

INRIA
Campus Universitaire de Beaulieu
35042 Rennes Cedex France

ABSTRACT

In this paper, we propose a novel inpainting algorithm combining the advantages of PDE-based schemes and examplar-based approaches. The proposed algorithm relies on the use of structure tensors to define the filling order priority and template matching. The structure tensors are computed in a hierarchic manner whereas the template matching is based on a K-nearest neighbor algorithm. The value K is adaptively set in function of the local texture information. Compared to two state of the art approaches, the proposed method provides more coherent results.

***Index Terms*—** inpainting, tensor, examplar-based.

1. INTRODUCTION

Inpainting methods play an important role in a wide range of applications. Removing text and advertisements (such as logos), removing undesired objects, noise reduction [1] and image reconstruction from incomplete data are the key applications of inpainting methods. There are two algorithm categories: PDE (Partial Derivative Equation)-based schemes [2] and examplar-based schemes [3]. The former uses diffusion schemes in order to propagate structures in a given direction. The drawback is the introduction of blur due to diffusion. The latter relies on the sampling and the copying of texture from the known parts of the picture.

In this paper, we propose a novel inpainting algorithm combining the advantages of both aforementioned methods. As in [3], the proposed method involves two steps: first, a filling order is defined to favor the propagation of structure in the isophote direction. Second, a template matching is performed in order to find the best candidates to fill in the hole. Compared to previous approaches, the main contributions are fourfold: the first one concerns the use of structure tensors to define the filling order instead of field gradients. Second is to use a hierarchical approach to be less dependent on the singularities of local orientations. Third is related to constraining the template matching to search for best candidates in the isophote directions. Fourth is a K-nearest neighbor approach to compute the final candidate. The number K depends on the local context centered on the patch to be filled in.

The paper is organized as follows. Section 2 describes the proposed method. Section 3 presents the performance of the method and a comparison with existing approaches. Section 4 concludes the paper.

2. ALGORITHM DESCRIPTION

The goal of the proposed approach is to fill in the unknown areas of a picture I by using a patch-based method. As in [3], the inpainting is carried out in two steps: (i) determining the filling order; (ii) propagating the texture. We use almost the same notations as in [3]. They are briefly summarized below:

- the input picture noted I. Let $I : \Omega \rightarrow \mathcal{R}^n$ be a vector-valued data set and I_i represents its i-th component;
- the region to be filled noted Ω;
- the source region noted ϕ, ($\phi = I - \Omega$);
- a square block noted ψ_p, centered at the point $\mathbf{p}$ located near the front line.

The differences between the approach in [3] and the proposed one are on one hand the use of structure tensors and on the other hand the use of a hierarchical approach. In the following, we first describe the algorithm for a unique level of hierarchical decomposition and then we describe how the hierarchical approach is used.

2.1. For one level of the hierarchical decomposition

As previously mentioned, the proposed approach follows the approach in [3] in the way that the inpainting is made in two steps. In a first step, a filling priority is computed for each patch to be filled. The second step consists in looking for the best candidate to fill in the unknown areas in decreasing order of priority. These two steps are described in the following sections.

2.1.1. Computing patch priorities

Given a patch ψ_p centered at the point $\mathbf{p}$ (unknown pixel) located near the front line, the filling order (also called priority) is defined as the product of two terms: $P(\mathbf{p}) = C(\mathbf{p})D(\mathbf{p})$.

The first term, called the confidence, is the same as in [3]. It is given by:

$$C(\mathbf{p}) = \frac{\sum_{q \in \psi_p \cap (I-\Omega)} C(\mathbf{q})}{|\psi_p|} \tag{1}$$

where $|\psi_p|$ is the area of ψ_p. This term is used to favor patches having the highest number of known pixels (At the first iteration, $C(\mathbf{p}) = 1 \; \forall \mathbf{p} \in \Omega$ and $C(\mathbf{p}) = 0 \; \forall \mathbf{p} \in I - \Omega$).

The second term, called the data term, is different from [3]. The definition of this term is inspired by PDE regularization methods acting on multivalued images [4]. The most efficient PDE-based schemes rely on the use of a structure tensor from which the local geometry can be computed. As the input is a multivalued image, the structure tensor, also called Di Zenzo matrix [5], is given by:

$$\mathbf{J} = \sum_{i=1}^{n} \nabla I_i \nabla I_i^T \tag{2}$$

(a) (b) (c) (d)

Fig. 1. (a) direction of the isophotes;(b) coherence norm: black areas correspond to areas for which there is no dominant direction; (c) Filling with the best candidate (K=1); (d) Filling with the best 10 candidates.

$\mathbf{J}$ is the sum of the scalar structure tensors $\nabla I_i \nabla I_i^T$ of each image channel I_i (R,G,B). The structure tensor gives information on orientation and magnitudes of structures of the image, as the gradient would do. However, as stated by Brox et al. [6], there are several advantages to use a structure tensor field rather than a gradient field. The tensor can be smoothed without cancellation effects : $\mathbf{J}_\sigma = \mathbf{J} * G_\sigma$ where $G_\sigma = \frac{1}{2\pi\sigma^2} exp(-\frac{x^2+y^2}{2\sigma^2})$, with standard deviation σ. In this paper, the standard deviation of the Gaussian distribution is equal to 1.0.
The Gaussian convolution of the structure tensor provides more coherent local vector geometry. This smoothing improves the robustness to noise and local orientation singularities. Another benefit of using a structure tensor is that a structure coherence indicator can be deduced from its eigenvalues. Based on the discrepancy of the eigenvalues, this kind of measure indicates the degree of anisotropy of a local region. The local vector geometry is computed from the structure tensor $\mathbf{J}_\sigma$. Its eigenvectors $\mathbf{v}_{1,2}$ ($\mathbf{v}_i \in R^n$) define an oriented orthogonal basis and its eigenvalues $\lambda_{1,2}$ define the amount of structure variation. $\mathbf{v}_1$ is the orientation with the highest fluctuations (orthogonal to the image contours), and $\mathbf{v}_2$ gives the preferred local orientation. This eigenvector (having the smallest eigenvalue) indicates the isophote orientation. A data term D is then defined as [7]:

$$D(\mathbf{p}) = \alpha + (1-\alpha) exp\left(-\frac{\eta}{(\lambda_1 - \lambda_2)^2}\right) \tag{3}$$

where η is a positive value and $\alpha \in [0,1]$ ($\eta = 8$ and $\alpha = 0.01$). On flat regions ($\lambda_1 \approx \lambda_2$), any direction is favored for the propagation (isotropic filling order). When $\lambda_1 >> \lambda_2$ indicating the presence of a structure, the data term is important.
Figure 1 shows the isophote directions (a) and the value of the coherence norm $(\frac{\lambda_1-\lambda_2}{\lambda_1+\lambda_2})^2$ (b). Black areas correspond to areas for which there is no dominant direction.

2.1.2. *Propagating texture and structure information*

Once the priority P has been computed for all unknown pixels $\mathbf{p}$ located near the front line, pixels are processed in decreasing order of priority. This filling order is called percentile priority-based concentric filling (PPCF). PPCF order is different from Criminisi's approach. Criminisi et al. [3] updated the priority term after filling a patch and systematically used the pixel having the highest priority. The advantage is to propagate the structure throughout the hole to fill. However, this advantage is in a number of cases a weakness. Indeed, the risk, especially when the hole to fill is rather big, is to propagate too much the image structures. The PPCF approach allows us to start filling by the $L\%$ pixels having the highest priority. The propagation of image structures in the isophote direction is still preserved but to a lesser extent than in [3]. Once the pixel having the highest priority is found, a template matching based on the sum of squared differences (SSD) is applied to find a plausible candidate. SSD is computed between this candidate (entirely contained in ϕ) and the already filled or known pixels of ψ_p. Finally, the best candidate is chosen by the following formula:

$$\psi_{\widehat{q}} = \arg \min_{\psi_q \in \mathcal{W}} d(\psi_{\widehat{p}}, \psi_q) \tag{4}$$

where $d(.,.)$ is the SSD. Note that a search window $\mathcal{W}$ centered on p is used to perform the matching.
Finding the best candidate is fundamental for different reasons. The filling process must ensure that there is a good matching between the known parts of ψ_p and a similar patch in ϕ in order to fill the unknown parts of ψ_p. The metric used to evaluate the similarity between patches is then important to propagate the texture and the structure in a coherent manner. Moreover, as the algorithm is iterative, the chosen candidate will influence significantly the result that will be obtained at the next iteration. An error leading to the apparition of a new structure can be propagated throughout the image. In order to improve the search for the best candidate, the previous strategy is modified as follows:

$$\psi_{\widehat{q}} = \arg \min_{\psi_q \in \phi} d(\psi_{\widehat{p}}, \psi_q) + (\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2})^2 \times f(p,q) \tag{5}$$

where the first term $d(.,.)$ is still the SSD and the second term is used to favor candidates in the isophote direction, if any. Indeed, the term $(\frac{\lambda_1-\lambda_2}{\lambda_1+\lambda_2})^2$ is a measure of the anisotropy at a given position (as explained in section 2.1.1). On flat areas, this term tends to 0. The function $f(p,q)$ is given by:

$$f(p,q) = \frac{1}{\epsilon + \frac{|\mathbf{v}_2 \cdot \mathbf{v}_{pq}|}{\|\mathbf{v}_{pq}\|}} \tag{6}$$

where $\mathbf{v}_{pq}$ is the vector between the centre p of patch ψ_p and the centre q of a candidate patch ψ_q. ϵ is a small constant value, set to 0.001. If the vector $\mathbf{v}_{pq}$ is not collinear to the isophote direction (line of constant intensity), this candidate is penalized by computing the scalar product $\mathbf{v}_2 \cdot \mathbf{v}_{pq}$. In the worst case (the two vectors are orthogonal), the penalization is equal to $1/\epsilon$. When the two directions are collinear, the function $f(p,q)$ tends to one.
A K nearest neighbour search algorithm is also used to compute the final candidate to improve the robustness. We follow Wexler et al.'s proposition [8] by taking into account that all candidate patches are not equally reliable (see equation 3 of [8]). An inpainting pixel $\widehat{c}$ is given by (c_i are the pixels of the selected candidates):

$$\widehat{c} = \frac{\sum_i s_i c_i}{\sum_i s_i} \tag{7}$$

(a) Mask (b) Proposed (c) Tschumperlé (d) Criminisi

(a) Mask (b) Proposed (c) Tschumperlé (d) Criminisi

(a) Mask (b) Proposed (c) Tschumperlé (d) Criminisi

Fig. 2. Comparison of the proposed approach with the approaches [2, 3].

where s_i is the similarity measure between the current patch and a candidate i (see equation 2 of [8]). The Euclidean distance is used. Most of the time, the number of candidates K is fixed. This solution is not well adapted. Indeed, on stochastic or fine textured regions, as soon as K is greater than one, the linear combination systematically induces blur. One solution to deal with that is to locally adapt the value K. In this approach we compute the variance σ_W^2 on the search window. K is given by the function $a + \frac{b}{1+\sigma_W^2/T}$ (in our implementation we use $a = 1$, $b = 9$ and $T = 300$. It means that we can use up to 10 candidates to fill in the holes). Figure 1 (c) and (b) shows the rendering of a fine texture with the best and the best ten candidates. For this example, good rendering quality is achieved by taking into account only the best candidate.

2.2. Hierarchical decomposition

Previous sections described the proposed approach for a given pyramid level. One limitation concerns the computation of the gradient ∇I used to define the structure tensor. Indeed, as the pixels belonging to the hole to fill are initialized to a given value (0 for instance), it is required to compute the gradient only on the known part of the patch ψ_p. This constraint can have a negative effect on the final quality. To overcome this limitation, a hierarchical decomposition is used in order to propagate throughout the pyramid levels an approximation of the structure tensor. A Gaussian pyramid is then built with successive low-pass filtering and downsampling by 2 in each dimension leading to nL levels. At the coarsest level $\mathcal{L}_0$, the algorithm described in the previous section is applied. For a next pyramid level $\mathcal{L}_n$, a linear combination between the structure tensors of level $\mathcal{L}_n$ and $\mathcal{L}_{n-1}$(after upsampling) is performed:

$$\mathbf{J}_h^{\mathcal{L}_n} = \nu \times \mathbf{J}^{\mathcal{L}_n} + (1-\nu) \times \uparrow 2(\mathbf{J}^{\mathcal{L}_{n-1}}) \quad (8)$$

where $\mathbf{J}_h$ is a structure tensor computed from a hierarchical approach. $\uparrow 2$ is the upsampling operator. In our implementation, ν is fixed and set to 0.6. This hierarchical approach makes the inpainting algorithm more robust to local orientation singularities. At the coarsest level, the local structure tensor is a good approximation of the local dominant direction. Propagating such information throughout the pyramid decreases the sensitivity to local orientation singularities and noise. By default, nL is set to 3.

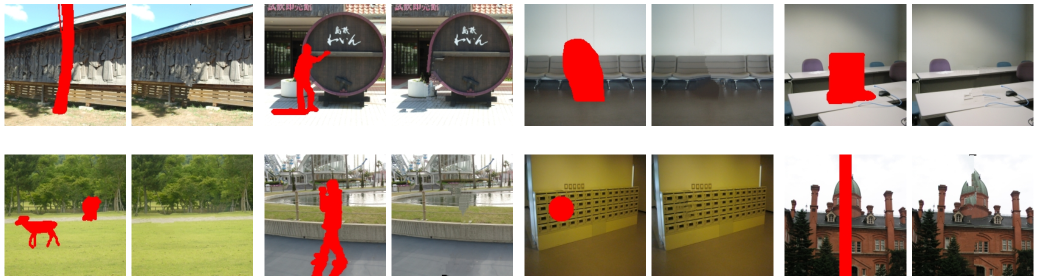
3. RESULTS

Figures 2 and 4 show the performance of the proposed method. A comparison with Criminisi et al.'s[1] and Tschumperlé et al.'s approach [2] is performed. Figure 2 depicts the results. The approach in [2] preserves quite well the images structures (except for the picture (c) of the second row) but the apparition of blur is annoying when filling large areas. On the first picture, results of the proposed and Criminisi's approach are similar. On the latter two, the proposed approach outperforms it. For instance, the roof as well as the steps of the third picture are much more natural than those obtained by Criminisi's method. The use of tensor and hierarchical approach brings an interesting visual gain[2]. For instance figure 3 shows the inpainted pictures with and without the hierarchical method. Figure 4 shows results on pictures belonging to Kawai et al.'s database [9][3]. Compared to previous assessment, these pictures have a smaller resolution (200 × 200 pixels) than those used previously (512 × 384).

[1]Matlab implementation available on http://www.cc.gatech.edu/~sooraj/inpainting/.

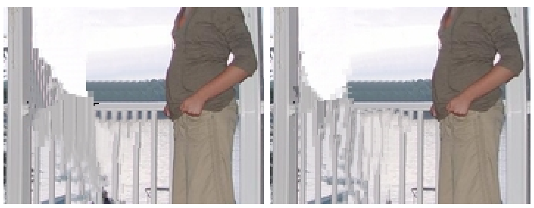
[2]see http://www.irisa.fr/temics/staff/lemeur/ to have more results.

[3]http://yokoya.naist.jp/research/inpainting/.



(a) Mask (b) Proposed (a) Mask (b) Proposed (a) Mask (b) Proposed (a) Mask (b) Proposed

Fig. 4. Results of the proposed approach on pictures proposed by [9].



(a) (b)

Fig. 3. Zoom on the picture of figure 2 (second row (b)) without the hierarchical approach (a) and with it (b).

The same setting than the previous one is used. As illustrated by the figure, the unknown regions have been coherently reconstructed. Except for the last picture, structures are well propagated without loss of texture information. A video clip showing how the proposed approach works is available on the author's web site.

4. CONCLUSION

In this paper, a novel inpainting approach is proposed. The novelties rely on the combination of advantages of PDE-based and examplar-based schemes. Structure tensors are computed in a hierarchic manner providing coherent information of local orientation as well as robustness to local orientation singularities. The filling process is based on a K-nearest neighbor approach for which the number of candidates used to fill in the hole is adaptively chosen in function of the local context. Compared to state-of-the-art methods, the proposed approach gives promising results. To enhance the robustness, local and dominant orientations (stemming from the structure tensors for the finest and the coarsest pyramid levels) might be used. Another point is related to the tuning of the proposed approach. For instance, the threshold T used to determine the number of candidate to use in the K-nearest neighbor approach is fundamental. Rather than using the local variance, it would be interesting to evaluate other criterion. Another improvement will focus on stochastic or inhomogeneous textures, for which repetitions of structure are absent. In this case, template matching fails to replicate this kind of texture in a coherent manner. Instead of using an exemplar-based method, it would be probably better to synthesize such texture by using stochastic-based texture models. It will be also required to compare the performance of the proposed approach to very recent methods such as [10].
Pictures used in this paper as well as software are available on `http://www.irisa.fr/temics/staff/lemeur/publi/2011_ICIP/1/index.php`.

5. ACKNOWLEDGEMENT

This work is supported by the French national program CONTINT through the project entitled PERSEE.

6. REFERENCES

[1] J. Weickert and H. Scharr, "An anisotropic diffusion algorithm with optimized rotation invariance," *Journal of Visual Communication and Image Representation*, vol. 13, no. 1-2, pp. 103–118, 2002.

[2] D. Tschumperlé and R. Deriche, "Vector-valued image regularization with pdes: a common framework for different applications," *IEEE Trans. on PAMI*, vol. 27, no. 4, pp. 506–517, April 2005.

[3] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by examplar(based image inpainting," *IEEE Trans. On Image Processing*, vol. 13, pp. 1200–1212, 2004.

[4] D. Tschumperlé, "Fast anisotropic smoothing of multi-valued images using curvature-preserving pde's," *Int. Jounral of Comp. Vision*, vol. 68, no. 1, pp. 65–82, 2006.

[5] S. Di Zenzo, "A note on the gradient of a multi-image," *Computer Vision, Graphics, and Image Processing*, vol. 33, pp. 116–125, 1986.

[6] T. Brox, J. Weickert, B. Burgeth, and P. Mrázk, "Nonlinear structure tensors," *Image and Vision Computing*, vol. 24, pp. 41–55, 2006.

[7] J. Weickert, "Coherence-enhancing diffusion filtering," *International Journal of Computer Vision*, vol. 32, pp. 111–127, 1999.

[8] Y. Wexler, E. Shechtman, and E. Irani, "Space-time completion of video," *IEEE Trans. On PAMI*, vol. 29, no. 3, pp. 463–476, 2007.

[9] N. Kawai, T. Sato, and N. Yokoya, "Image inpainting considering brightness change and spatial locality of textures and its evaluation," in *PSIVT2009*, 2009, pp. 271–282.

[10] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE TIP*, vol. 19, no. 5, pp. 1153–1165, 2010.